ROUGH SETS APPLICATION FOR STUDENTS CLASSIFICATION BASED ON PERCEPTUAL DATA

Asheesh Kumar*, Naresh Rameshrao Pimplikar*, Apurva Mohan Gupta*

VIT University Vellore -14 India

Abstract - Now days, Artificial Intelligence is used across the different range of areas including Industrial Automation, Robotics, Logistic and Transport, Avionics, Arts and Media, Virtualization and it also will prominent in future. Different intelligent techniques include the ANN, Fuzzy logic and Fuzzy sets, Genetic algorithms and other methods as soft computing. The rough set theory is an import concept that can also bring a new horizon in this area. In this paper the theme is discussion about the rough set theory and some potential industrial applications drafted. Rough set theory deals with the inadequate data and gives the approximate output. This paper discussion carries on the applications of AI, conceptual overview of rough set theory and its application with the educational data classification.

Index Terms - Rough set theory, Educational data classification, Applications of rough set theory.

I. INTRODUCTION

Rough set theory is a mathematical formulation for the data analysis. The rough se theory concept works on the grouping or making the similar clusters that holds entities that are undetectable for some futures such as temperature, motivational factor of human, socialized index etc., these features are the basic augmenting facts of knowledge that tells about the reality, and used for discovering a secret pattern in a given data set. The introduction of the rough set theory can be discovered in Pal and Skowron (1999).

For data analysis different methods are used that has some intercept with the rough set theory such as fuzzy sets, incomplete data or evidence theory, clustering and ANN but rough set theory has its own access that support it as an autonomous area. Rough set theory can be implemented with the real life applications with the various areas of science and technology. It is used for the decision making and approximate calculation of the data sets. In the real world, the rough set theory can be used for the knowledge abstraction and after that decision rule can be applied for getting the result. The rough set theory become very popular for scientific data analysis and controls the automation in the different types of industrial applications. This paper describes about the decision making on educational data for the classification purpose. After the classification the categorization or grouping of the students provide essential information about a group of students. This paper carries the discussion on use of the various perceptual values of students and makes a decision on these input values.

II. LITRATURE REVIEW

In past few years rough set theory had gain much importance in various industrial, cognitive and database applications. In rough set theory lots of work is done by various prominent authors and scholors. Plonka, Peters, Ziarko had done various intelligent control system design using rough set theory. Machine fault tolerance is implemented by Nowicki etc. Rough set theory approach is implemented in the decision system, drug design, biomedical engineering, medical imaging (Zak & Stefanowski).

III. RST approach for categorization of students and relationship with KRR

Rough set theory approach for knowledge extraction from student data for classification is related to the KRR. KRR is the study of knowledge representation using symbolically and make the reasoning based on the logic. This paper, first educational data sets are represented using symbols in the data table format that is refer as representation and then apply the rough set theory method (rules generated by RST)for reasoning. So the KRR and data representation, reasoning, extraction and making the decision rules based on the RST.

IV. USE OF ROUGH SET THEORY AND INTELIGENT INDUSTRIAL APPLICATIONS

Now days AI is the basic concept to control the automation in industry and it is also used for making the decision in the critical situation and in near future it plays important role for controlling the production industry, auto pilot, drug research, detecting the disease in human being. Rough set theory is applicable with various fields that can respond automatically such as power control in steel plant. Rough set theory is used for the various areas that are discussed below.

1) Manufacturing and Material engineering: The rough set theory is used in the material engineering for detecting the properties of material that essentially define use of that material in various applications. It is used in manufacturing engineering for controlling the manufacturing process such as power production control and heavy industry. The initiative work in this field is done by the Jackson etc. all (1994).
2) Intelligent control: Rough set theory used for controlling the various industrial applications such as linear system in the aerodynamics modeling. But as compared to the other techniques such as fuzzy logic and ANN provides highly reasonable results. And rough set theory also provides a good result in many cases. It can be used for controlling Steal kiln control, Furnace temperature control etc. In this area of rough set application is done by (Lin 1997, Ziarko 1992 and Mrozek 1992).
3) Decision support system: The rough set theory approach for making the decision used for different variety of industrial applications.It can be also used for making a decision in financial services and other managerial and business levels; the work done in this field is (Stepaniuk 1996, Ziarko et al. 1993).
4) Detecting the fault in the mechanical engineering: Rough set theory technique is used for detecting the fault occur at the different component of a mechanical device and diagnosis; the work done in this field is by (Slowinski et al. 1996, Nowicki et al 1992)
5) Neural networks: Neural networks are used for making a decision and control the industrial components. Neural network and rough set theory combines together and provide a new horizon in the field of industrial control as compared to alone neural network approach; the work done in this field is by (Mitra et al 1996, Lingras 1996)
6) Other possible domain: So the other discipline are also there where the rough set theory can also be applied and provide a good results such as data mining, modeling and design, telecommunication and expert systems.

Rough set approach has many advantages that make understand to use this approach in industrial and decision making process such as searching a data pattern in dataset, making a decision rules, result can be understood easily also can be applied quantitative data etc. So the rough set theory display have its own advantage and can be applied to the various fields of industrial application process control and it can also be combined with the other existing techniques such as fuzzy logic and ANN and provides a better result.

ISSN: 2278 – 1323

V. ROUGH SET THEORY: BASIC CONCEPTS

Data is generally represented in the table format called data table, data table is also referred as database or attribute value table. In the data table matrix representation objects are represented or labeled with row and attributes are represented with the columns. Entries entered into the matrix are called attribute values. A typical attribute value table example are shown below in the Table number 1,

Students	A	B	C	D
1	honor	good	no	D.M.A.Y.
2	medium	good	no	D.M.A.N.
3	medium	good	no	D.M.A.Y.
4	below average	average	no	D.M.A.N.
5	medium	average	yes	D.M.A.N.
6	honor	average	yes	D.M.A.Y.

Table No. 1: Information Table example

In the information table represents six students that have core attributes:

A – Marks of the students
B – Social acceptance factor
C – Achievements
D – Decision Making Ability

Now we are concerned about the features affiliated with the decision making ability D.M.A.Y. /D.M.A.N. of students. This question cannot be resolved as unique since a data table provides the inconsistent data such as student 2 and 3 has the same characteristics but student 2 has D.M.A.N. whereas student 3 has DMAY. The condition is pictured in Fig. 1

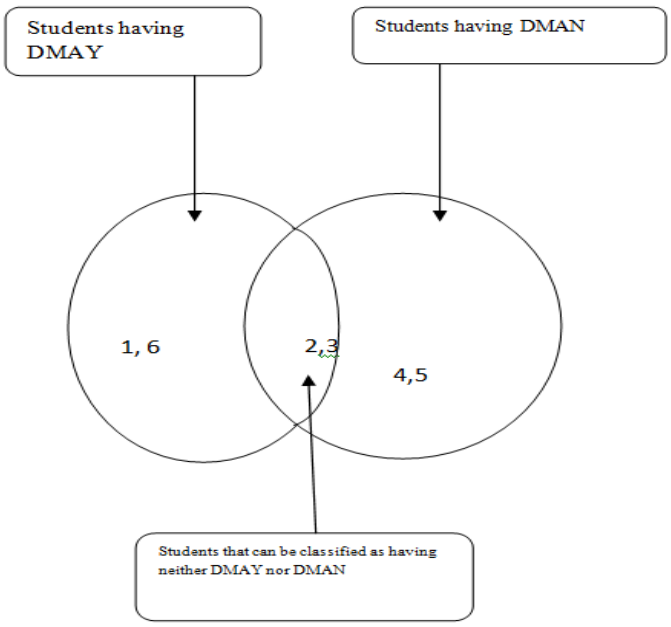

Fig.1 Student's Classification

Under the normal condition, this form of inconsistency or fuzzy characterization of students is adopted. The inconsistency can also be removed by using more attributes in the database. But in the rough set theory we apply a different approach to solve this kind of inconsistence problem. In rough set theory it is tried to conserve entire data and find out other directions, and then analyze what actual data is assured about. So in the rough set theory a concept of lower and upper approximation of sets, that is actual topological interior and closure operations. So, the different classes of the students are identified as given below:

- The set {1, 3, 6} of all students having decision making ability yes (DMAY).
- The set {1, 6} of all students certainly show the decision making ability yes and this is the lower approximation of the previous set.
- The set {1, 2, 3, 6} of all students possibly showing the decision making ability yes and it is the upper approximation of the set {1, 3, 6}.
- The set {2, 3} of all students may be grouped as neither DMAY nor DMAN, and it is the boundary region of the set {1, 3, 6}.
- The set {2, 4, 5} of all students deliver the DMAN.
- The set {4, 5} of all students certainly deliver DMAN, this is lower approximation of the set {2, 4, 5}

- The set {2, 3, 4, 5} of all students possibly deliver DMAN, this is the upper approximation of set {2, 4, 5}
- The set {2, 3} of all students may be grouped as delivered neither DMAY nor DMAN, this is the boundary result of the set {2, 4, 5}

In the above example, data table is divided in the group that deliver the similar attribute values, or showing the same concept delivering in the terms of attribute values. So each group is exposing the same characteristics of the objects that are used for making the block, which is understood as basic granules of knowledge. These are known as the elementary sets or concepts, and conceive as the basic building blocks of knowledge about reality, are interested in. Compound concepts are the combination of the elementary concept. The crisp set is defined as the union of elementary sets and other set is known as the rough set. Lower and upper approximations of a set X are the combination of two crisp sets. Lower approximation of a set is the set of all elements that certainly belongs to X and upper approximation is the set of elements that possibly belong to the X. The different of upper and lower approximation of X is the boundary region. So here rough set is defined as non-empty boundary region, otherwise set is crisp. Approximation is the basic operation for dealing with uncertain and undefined data in rough set theory.

Now we consider the two finite non empty sets U and A. A denotes the set of attributes and U denotes the universe. For each attribute a ∈ A, and value of attribute a belongs to the set V_a, this is known as the domain of A. And S is known as the database and denoted by the (U, A). Now we consider a subset B of A maps a binary relation I(B) on U, that is known as the indiscernibility, shows as:

If (x, y)∈I(B) iff a(x) = a(y) for every a ∈ A, where a(x) denotes the value of attribute a for element x. If (x, y) ∈I(B) then x , y are B-indiscernible. Equivalence classes of the relation I(B) are known as the B-elementary concepts:

$$B_*(X) = \bigcup_{x \in U} \{B(x) \in U : B(x) \subseteq X\}$$

$$B^*(X) = \bigcup_{x \in U} \{B(x) \in U : B(x) \cap X \neq \emptyset$$

B_* is referred as B lower and B upper approximation of X.

$$BN_B(X) = B^*(X) - B_*(X)$$

And this is known as B boundary of X

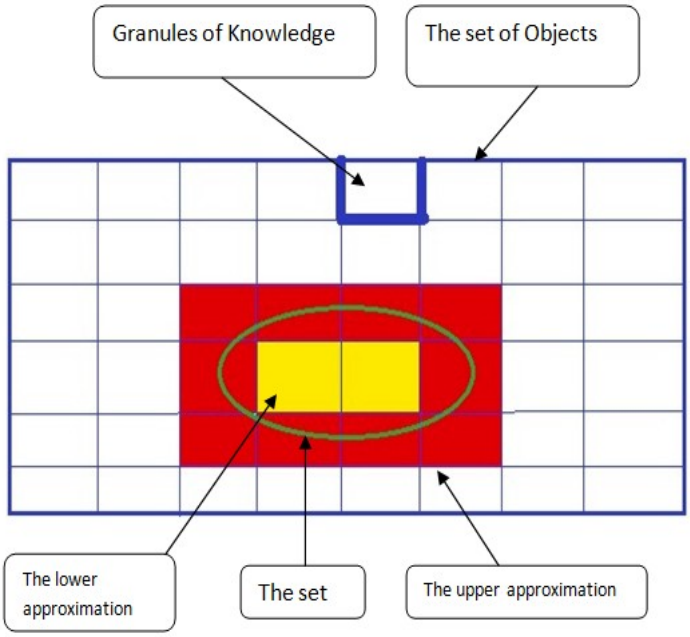

Fig.2. Pictorial representation of Rough set theory

A. MEMBERSHIP FUNCTION

The rough membership function, is specified as

$$\mu_X^B(x) = \frac{card(X \cap B(x))}{card(B(x))}.$$

$$0 \leq \mu_X^B(x) \leq 1.$$

$\mu_x(x)$ is conditional probability and also can be understood as a degree of certainty, x belongs to X. The rough membership function specifies the approximations values and boundary region:

$$B_*(X) = \{x \in U : \mu_X^B(x) = 1\},$$
$$B^*(X) = \{x \in U : \mu_X^B(x) > 0\},$$
$$BN_B(X) = \{x \in U : 0 < \mu_X^B(x) < 1\}.$$

ISSN: 2278 – 1323

B. ATTRIBUTE DEPENDENCY

The database showing the condition and decision attributes. Decision attributes totally depends on the conditional attributes, C used for the notation condition and D for the decision and denoted by C⇒ D, D is totally dependent on the values of the C that means there is a functional dependency between the decision and condition. The partial dependency is defined as the only few values of the decision is depends on the values of condition. And it is denoted by the C $\Rightarrow_k$D, k is used for the as notation degree of dependency and it always lies between 0 to 1.

$$k = \gamma(C,D) = \sum_{X \in U/D} \frac{card(C_*(X))}{card(U)}.$$

As values given in the Table no.1, the degree of dependency, set of attributes {A, B, C} and the D is 2/3.

C. ATTRIBUTE REDUCTION

In the rough set theory we pay attention for preserving the properties but removal of the some values also possible without affecting a database property. So reduction of the attributes in the database is possible without changing the meaning of database. C is the condition and D is the decision attributes, C, D ⊆A. Now C` ⊆ C is a D-reduct of C, if C` is a minimal subset of C:

$$\gamma(C,D) = \gamma(C',D).$$

So reduction of the conditional attributes of the database in a manner that the dependency between the attributes of condition and decision are conserved. In the table no. 1 {A,B,C} attributes has {A,B} and {A,C} reducts.

D. DECISION RULES

Dependency of the condition and decision are represented in the decision rule as „"if …Then". Φ → Ψ represents the decision rule, Φ denotes the condition and Ψdenotes the decision. In the rough set theory connectives can be used to make a formula. Such that in table no.1 decision rule: if (A, honor) and (B, good) then (D, DMAY).

In decision rule Ψ is true, ifΦ is true in the given S, so the probability $\pi_S(\Phi)$,is certainty factor:

$$\pi_S(\Psi \mid \Phi) = \frac{card(\| \Phi \wedge \Psi \|_S)}{card(\| \Phi \|_S)}$$

$\|\Phi\|_S$ use for notation the all objects in S having properties showed by the Φ.

The coverage factor of decision rule is defined as

$$\pi_S(\Phi \mid \Psi) = \frac{card(\| \Phi \wedge \Psi \|_S)}{card(\| \Psi \|_S)}$$

This is the conditional probability that Φis true then Ψ is true with the probability $\pi S(\Psi)$.

In rough theory, two type of decision rules:

- If the value $\pi_s(\Psi \mid \Phi) = 1$ **then the** value of decision is totally depend on the condition Φ →Ψ defined as the certain or sure. e.g. If (A, honor) and (B, good) and (C, no) then (D, DMAY) is sure.
- If Φ →Ψ is the possible, if the value of $\pi_s(\Psi \mid \Phi) < 1$. E.g. If(A, medium) and (B,good) and (C, no) then (D, DMAN) is possible.

Some approximations also can be expressed with the help of the minimal set of DR:

1) If (A, honor) then (D, DMAY)
2) If (A, medium) and (B, good) then (D, DMAY)
3) If (A, below average) or ((A, medium) and (B, average)) then (D, DAMN)
4) If (A, medium) and (B, good) then (D, DMAN)

The above rules can also be converted in the inverse, e.g. If decision is true then condition is true by the rule of inverse. Such as rule no. 1 is converted in inverse, if (D, DMAY) then (A, honor).

Students	A	B	C	D	Certainty factor	Coverage factor
1	honor	good	no	D.M.A.Y.	1	1/3
2	medium	good	no	D.M.A.N.	1/2	1/3
3	medium	good	no	D.M.A.Y.	1/2	1/3
4	below average	average	no	D.M.A.N.	1	1/3
5	medium	average	yes	D.M.A.N.	1	1/3
6	honor	average	yes	D.M.A.Y.	1	1/3

Table no. 2: Certainty factor and coverage factor calculation based on Table no.1

VI. ROUGH SET THEORY IMPLEMENTATION ON STUDENTS DATABASE FOR CLASSIFICATION

In this section, talk about the rough set theory approach implementation for the students classification based on the various perceptual properties such as achievements, behavioral ability etc. The Table no.1 describes the various inputs and outputs in the form of matrix that called data table. The input values are conditions and output values are decisions. In table no.1 the input or conditions are given below:

A – Marks of the students
B – Social acceptance factor
C – Achievements
D – Decision Making Ability

Range of the parameters of above values is:

V_A- {1, 2, 3, 4}, where 1=[60-70], 2=[70-80], 3=[80-90], 4=[90-100]

V_A- {1, 2, 3, 4}, where 1= [60-70], 2=[70-80], 3=[80-90], 4=[90-100]

V_B - {1, 2, 3}, where 1= un-socialized person (isolated), 2=Normal socialized, 3= Highly Socialized

V_C - {1, 2, 3}, where 1= Extra-curricular achievement, 2=Academic achievement, 3= both achievements

V_D - {1, 2}, where 1= Moderate decision making, 2= Quick decision making.

These conditional attributes are used for making the decision about the overall performance index of the students. Based on the performance index, a group of student formed, similar characteristics student fall in the same cluster.

E – PerformanceIndex

V_E - {2, 3}, where 2= Normal performance, 3= Goodperformance.

In Table no.3 decision made by the observer during the one academic session is given below.

Applying the rough set technique from Table no.3, rules are generated and given below:

1) If (A, 3) then (E, 3)
2) If (A, 3) or (B, 3) then (E, 3)
3) If (A, 4) then (E, 3)
4) If (A, 3) and ((B, 3) or (D, 2)) then (E, 3)
5) If (A, 3) and ((A, 4) or (B, 3)) then (E, 3)
6) If (A, 2) or (D, 1) then (E, 2)
7) If (A, 2) and ((C, 2) or (D, 1)) then (E, 2)

Rule generation see Mrozek 1989.

So based on the rules generated by RST, the new decision values are almost same as observer values calculated. Thus, the rough set theory produces the approximate same result as observer generated it by using any other soft computing techniques. So, rough set theory can be used for the classification of the students based on the perceptual values

ISSN: 2278 – 1323

Students No.	Conditions				Decisions	Students No.	Conditions				Decisions
	A	B	C	D	E		A	B	C	D	E
1	3	3	1	2	3	11	2	1	2	1	2
2	1	2	1	1	2	12	1	1	1	1	2
3	2	2	2	2	2	13	2	2	3	2	2
4	4	3	2	2	3	14	4	2	3	2	3
5	1	3	1	2	2	15	1	2	1	2	2
6	3	3	1	2	3	16	3	1	1	1	2
7	2	2	1	1	2	17	2	2	3	2	2
8	1	1	1	1	2	18	4	3	2	2	3
9	4	3	2	2	3	19	2	2	1	1	2
10	3	3	1	2	3	20	3	2	3	3	3

Table No. 3 Protocol of Observer Decisions

Students No.	Conditions				Decisions	New Decision	Students No.	Conditions				Decisions	New Des.
	A	B	C	D	E			A	B	C	D	E	
1	3	3	1	2	3	3	11	2	1	2	1	2	2
2	1	2	1	1	2	2	12	1	1	1	1	2	2
3	2	2	2	2	2	2	13	2	2	3	2	2	3
4	4	3	2	2	3	3	14	4	2	3	2	3	3
5	1	3	1	2	2	2	15	1	2	1	2	2	2
6	3	3	1	2	3	3	16	3	1	1	1	2	3
7	2	2	1	1	2	2	17	2	2	3	2	2	2
8	1	1	1	1	2	2	18	4	3	2	2	3	3
9	4	3	2	2	3	3	19	2	2	1	1	2	2
10	3	3	1	2	3	3	20	3	2	3	3	3	3

Table no.4: Observer decision vs. Rough set rule decisions

VII. CONCLUSION AND FUTURE RESEARCH

Considering the various soft computing techniques likewise fuzzy sets, ANN, RST is also gives the reasonable results, so it is also considerable nominee. The leading industrial areas of the rough set theory application are drug research, fault diagnosis in machines, bioinformatics, medical data analysis and decision support. Rough set theory has superior characteristics like finding the relationship and data pattern and that will not use any statistics method and it allows qualitative and quantitative information, and also uses a direct pointed explanation of the results. In this paper, discussed application based on the educational data classification of the students in the group is very easy. Rules generated by the given perceptual data give the almost same results as calculated manually.

The future research in the rough set theory application in this field is, based on the rule generated information retrieval in a large database. Another future research scope is making an agent base system that can provide the assistance to the appropriate candidate based on the group in which candidate is falling.

REFERENCES

[1] Doreswamy, Hemanth, Vastrad, Application of rough sets for Engineering Materials Classification, IEEE 2010 ICINA, Vol 6, Page No. 401 – 406.

[2] Busse, Rough set theory with applications to data mining,Springer 2008 ICE, March Vol 3, Page No. 101 – 124.

[3] ZdzislawPawlak,Rough set theory and its applications, Journal of telecommunication and information technology, March 2002, Page No. 7 – 12.

[4]Hassanien, Ali, Rough set approach for generation of classification rules of breast cancer data, Informatuca, 2004, Vol 15, No. 1 Page No. 23 – 38.

[5] Kazuhisa Chiba, Efficiency of rough set theory as decision making optimization problem, EADMFDOAC 2011, Page No. 1 – 10.

[6]Sabu M K, Raju G, Rule induction using rough set theory and application in aggriculture, IEEE ICCCET 2011, 18 and 19 March 2011, Page No. 45 – 49.

[7]Baiting, Xijun, Qingshuang, Rough sets based hybrid intelligent fault diagnosis for precision test turntable, IEEE 2009, 9140HT, Page No. 1 – 5.

[8] Wang, Wang, Zhang, Yan, Decision rule extraction method based on rough set theory and fuzzy set theory. IEEE 2005, Fourth ICMLC, Page No. 2212 – 2216.

[9] Wang, Wang, Zhang, Yan, Decision rule extraction method based on rough set theory and fuzzy set theory. IEEE 2005, Fourth ICMLC, Page No. 2212 – 2216.

[10] Herawan, Ghazali, Yanto and Deris, Rough set approach for categorical data clustering, Springer 2010, March 2010, IJDTA,Vol 3, Page No. 33 – 52.

[11] Mrózek, A.: Rough sets and dependency analysis among attributes in computer implementations of expert's inference models. International Journal of Man-Machine Studies 30 (1989) 457-473.

Asheesh Kumar is M. Tech (CSE) Student at VIT University Vellore TamilNadu India. He did his B.Tech (IT) from IIMT Engineering College, Meerut, UP. His area of Interest is Algorithms Analysis, Theory of Computation, Web Services and Cloud Computing

Naresh Rameshrao Pimplikar is M. Tech (CSE) Student at VIT University Vellore TamilNadu India. He did his BE (IT) from Priyadarshini College of Engineering, Nagpur, Maharashtra. His area of Interest is Database Management System, Operating System, and Computer Graphics.

Apurva Mohan Gupta is M. Tech (CSE) Student at VIT University Vellore TamilNadu India. He did his BE (CSE) from Global Institute of Technology, Jaipur, Rajasthan. His area of Interest is Data Structure, Computer Architecture and Computer Graphics.

.

ISSN: 2278 – 1323